

Random nonlinear model with missing responses

Gabriela CIUPERCA¹

Université de Lyon, Université Lyon 1, UMR 5208, blvd du 11 novembre 1918, F - 69622 Villeurbanne Cedex, France

Abstract

A nonlinear model with response variable missing at random is studied. In order to improve the coverage accuracy, the empirical likelihood ratio (EL) method is considered. The asymptotic distribution of EL statistic and also of its approximation is χ^2 if the parameters are estimated using least squares(LS) or least absolute deviation(LAD) method on complete data. When the response are reconstituted using a semiparametric method, the empirical log-likelihood associated on imputed data is also asymptotically χ^2 . The Wilk's theorem for EL for parameter on response variable is also satisfied. It is shown via Monte Carlo simulations that the EL methods outperform the normal approximation based method in terms of coverage probability up to and including on the reconstituted data. The advantages of the proposed method are exemplified on the real data.

keywords: random nonlinear model; response missing at random; empirical likelihood; semi-parametric estimation;

1 Introduction

It is very common in practice, for a model, to measure the regressors(covariates) but for varied reasons it is sometimes impossible to have all values of the response variable. The most widely used idea is to remove of model the observations with missing data. An alternative solution is to consider empirical likelihood(EL) method which is a powerful nonparametric method for constructing confidence regions of parameters.

To our knowledge, previous theoretical and numerical investigations in literature have focused for model with missing response only in the linear case. The EL method, proposed by Owen (1990) does not need the asymptotic

¹ Tel:(33)0472431690; Fax:(33)0472431687

E-mail address: Gabriela.Ciuperca@univ-lyon1.fr

variance of estimator and it outperforms the normal approximation method in term of coverage probability for linear models. Wang and Rao (2002) develops EL inferences for the mean of a response variable under regression imputation of missing responses for a linear regression model and random covariates. Qin et al. (2009) construct an EL statistic on parameter when regressors are deterministic and Xue (2009a) if regressors are random, based on least squares(LS) method for linear model $Y = \mathbf{X}^t\boldsymbol{\beta} + \varepsilon$. Sun and Wang (2009) consider the general linear model $Y = \mathbf{H}(\mathbf{X})^t\boldsymbol{\beta} + \varepsilon$ with $\mathbf{H}(\mathbf{x})$ a known vector function and investigate a hypothesis test on the response variable. These last three papers impose the condition that the conditional expectation of error ε with respect to covariate \mathbf{X} is zero. If this hypothesis is not satisfied, in order to reconstitute the response variable the least absolute deviations(LAD) method can be used. One advantage of least absolute deviations estimation is that it does not require any moment condition on the errors to obtain asymptotic normality.

It is well known also that one outlier may cause a large error in a least squares estimator. This occurs in the case of fatter tail distributions of the error term. On the other hand, as Bai (1998) and Kim and Choi (1995) indicate it, for heavy tailed distributions the LAD estimator is more efficient than LS estimator.

Concerning the LAD estimator in a complete nonlinear model we can refer to following papers: Oberhofer (1982) shows conditions for its consistency, Weiss (1991) proves that this estimator is consistent and asymptotically normal in a dynamic nonlinear model with neither independent nor identically distributed errors. Ciuperca (2010) gives the convergence rate v_n , where (v_n) is a monotone positive sequence such that $v_n \rightarrow 0$ and $nv_n^2 \rightarrow \infty$ for $n \rightarrow \infty$. It is well known that confidence regions based on the asymptotic normality could encounter large coverage errors in small sizes or if the error distribution has outliers. The LAD technique was already used in censored median linear regression model with missing data: see e.g. Zhao and Chen (2008).

For other relevant papers (not exhaustive list) on EL method for missing data in a linear models see Wang and Sun (2007), Sun et al. (2009), Wang et al. (2004), Yang et al. (2009), Liang et al. (2009).

In this paper, for a nonlinear random model $Y = f(\mathbf{X}; \boldsymbol{\beta}) + \varepsilon$ with missing responses, some empirical likelihood ratios are constructed by using complete-case or imputed values. The nonparametric version of Wilks' theorem is proved for two cases: the parameters are estimated by least squares and least absolute deviations on complete data. The limiting distribution of EL statistic is χ^2 , results which can be used to construct confidence region on parameter. In order to complete data, the value imputed of missing value of variable response is obtained by generalising Xue (2009a) idea for linear model using a semiparametric technique: the parameters regression are estimated by LS method and missing probability by a nonparametric method. We show that the empirical log-likelihood ratio on parameter based on the improved data is asymptotically chi-squared. The numerical simulations proves that EL methods outperforms

the normal approximation method in terms of coverage probability up to and including on the reconstituted data. If the distribution of the errors presents outliers, the LAD method gives generally best results than LS method on coverage probability and on parameter estimators efficiency. In addition to that, if the expectation of error does not exist, as it is the case for Cauchy distribution, the normal approximation of LS estimator can not be satisfied. On the real data we obtain also that our semiparametric method gives more precise results to reconstitute the response variable than classic parametric LS method.

The rest of this paper is organised as follows. In Section 2 we introduce model, assumptions and some notations. In Section 3 the Wilks' theorem for EL statistic and also for its approximation is given, when the parameters are estimated using LS or LAD method on complete data. The LS case is developed in Section 4, a reconstituted value for the response variable is introduced and asymptotic distribution of EL for response variable is obtained. Section 5 illustrates by simulation results that EL methods for nonlinear random model outperform the normal method and give very competitive coverage probabilities. An application to the real data is presented in Section 6. Finally, Section 7 contains the proofs of the lemmas and of the theorems.

2 Notations and model

Let us consider following random nonlinear model:

$$Y_i = f(\mathbf{X}_i, \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where $(\varepsilon_i, \mathbf{X}_i)$ is a sequence of continuous independent random vectors with the same joint distribution as $(\varepsilon, \mathbf{X})$ and $\boldsymbol{\beta}$ is a $d \times 1$ vector of unknown regression parameters. More precisely, ε is a random variable and \mathbf{X} a $p \times 1$ random vector of covariates. Let $\boldsymbol{\beta}^0$ denote the true (unknown) of the parameter $\boldsymbol{\beta}$.

With regard to the random variable ε we make one of the suppositions:

(H1) $\mathbb{E}[\varepsilon_i | \mathbf{X}_i] = 0$ and $\mathbb{E}[\varepsilon_i^2 | \mathbf{X}_i] < \infty, \forall i = 1, \dots, n$.

(H1bis) $\mathbb{E}[\text{sign}(\varepsilon_i) | \mathbf{X}_i] = 0$ and ε_i have continuous density $e(t)$ satisfying $e(0) > 0$.

These conditions are essential for the consistency and asymptotic normality of the LS, respectively LAD estimator.

In following we use notation $\dot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta}) = \partial f(\mathbf{x}, \boldsymbol{\beta}) / \partial \boldsymbol{\beta}$, $\ddot{\mathbf{f}}(\mathbf{x}, \boldsymbol{\beta}) = \partial^2 f(\mathbf{x}, \boldsymbol{\beta}) / \partial \boldsymbol{\beta}^2$. For a vector, let us denote $\|\cdot\|$ the Euclidean norm.

Regression function $f : \Upsilon \times \Gamma \rightarrow \mathbb{R}$, with $\Upsilon \subseteq \mathbb{R}^p$, $\Gamma \subseteq \mathbb{R}^d$ and random vector \mathbf{X} satisfy the conditions:

(H2) for all $\mathbf{x} \in \Upsilon$ and for $\boldsymbol{\beta} \in \Gamma$, the function $f(\mathbf{x}, \boldsymbol{\beta})$ is twice differentiable

in β and continuous on Υ .

(H3) $\mathbb{P}[\|\mathbf{X}\| > M_n] = o(n^{-1/2})$ for any positive sequence $M_n \rightarrow \infty$ as $n \rightarrow \infty$.

(H4) $\|\mathbf{f}(\mathbf{x}, \beta)\|$, $\|\dot{\mathbf{f}}(\mathbf{x}, \beta)\|$ are bounded for any $\mathbf{x} \in \Upsilon$ and β in a neighborhood of β^0 .

Sets Υ and Γ are compacts. This, assumptions (H2) and (H4) are commonly used in nonlinear modelisation and are necessary for the consistency and for the asymptotic normality of the LS or LAD estimator. Assume also that the model is identifiable: if $f(\mathbf{X}; \beta) = f(\mathbf{X}; \beta^*)$ with probability one, then $\beta = \beta^*$.

For model (1), all the \mathbf{X}_i 's are observed, in exchange response variable Y_i can be missing. Let be the sequence of random variables $(\delta_i)_{1 \leq i \leq n}$ defined by: $\delta_i = 0$ if Y_i is missing and $\delta_i = 1$ if Y_i is observed. We suppose that Y_i is missing at random(MAR): $\mathbb{P}[\delta_i = 1 | \mathbf{X}_i, Y_i] = \mathbb{P}[\delta_i = 1 | \mathbf{X}_i]$, for all $i = 1, \dots, n$. Consider the selection probability function: $\pi(\mathbf{x}) = \mathbb{P}[\delta = 1 | \mathbf{X} = \mathbf{x}] > 0$, $\forall \mathbf{x} \in \Upsilon$. The supposition $\pi(\mathbf{x}) > 0$ is a common assumption in the literature. The parameter β is estimated on the completely observed data by two methods: least squares(LS) method:

$$\hat{\beta}_{n,LS} = \arg \min_{\beta} \sum_{i=1}^n \delta_i [Y_i - f(\mathbf{X}_i, \beta)]^2 \quad (2)$$

and least absolute deviations(LAD) method:

$$\hat{\beta}_{n,LAD} = \arg \min_{\beta} \sum_{i=1}^n \delta_i |Y_i - f(\mathbf{X}_i, \beta)|. \quad (3)$$

To build the EL statistic, let us consider following functions, for $i = 1, \dots, n$:

$$\mathbf{g}_{i,LS}(\beta) = \delta_i [Y_i - f(\mathbf{X}_i, \beta)] \dot{\mathbf{f}}(\mathbf{X}_i, \beta),$$

$$\mathbf{g}_{i,LAD}(\beta) = \delta_i \text{sign}(Y_i - f(\mathbf{X}_i, \beta)) \dot{\mathbf{f}}(\mathbf{X}_i, \beta)$$

and let be also:

$$\mathbf{g}(\beta) = \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\beta),$$

with $\mathbf{g}_i(\beta)$ either $\mathbf{g}_{i,LS}(\beta)$ or $\mathbf{g}_{i,LAD}(\beta)$. The two estimators are a solution of the system of equations: $\mathbf{g}(\beta) = \mathbf{0}$. Under (H1), respectively (H1bis), we have: $E[\mathbf{g}_i(\beta)] = \mathbf{0}$ for $\mathbf{g}_i = \mathbf{g}_{i,LS}$, respectively $\mathbf{g}_{i,LAD}$. The joint distribution of the errors ε_i and of \mathbf{X}_i is unknown, but conditional mean, respectively median, of ε_i is zero.

The empirical likelihood for β with complete-case data can be defined as (see

Owen (1990)):

$$L_n(\boldsymbol{\beta}) = \sup_{(p_1, \dots, p_n) \in (0,1)^n} \left\{ \prod_{i=1}^n p_i; \quad \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{0} \right\}. \quad (4)$$

Without constraint $\sum_{i=1}^n p_i \mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{0}$, the maximum of $\prod_{i=1}^n p_i$ is attained for $p_i = n^{-1}$. Then, the profile empirical likelihood ratio for $\boldsymbol{\beta}$ with complete-case has the form:

$$R_n(\boldsymbol{\beta}) = \sup_{(p_1, \dots, p_n) \in (0,1)^n} \left\{ \prod_{i=1}^n np_i; \quad \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{0} \right\}. \quad (5)$$

The empirical log-likelihood ratio statistic evaluated at $\boldsymbol{\beta}$ is:

$$l_n(\boldsymbol{\beta}) = -2 \sup_{(p_1, \dots, p_n) \in (0,1)^n} \left\{ \sum_{i=1}^n \log(np_i); \quad \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \mathbf{g}_i(\boldsymbol{\beta}) = \mathbf{0} \right\}. \quad (6)$$

Let $\boldsymbol{\lambda} \in \mathbb{R}^d$ be, a Lagrange multiplier. Then, $R_n(\boldsymbol{\beta})$ is maximised for $p_i = n^{-1}[1 + \boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})]^{-1}$, where $\boldsymbol{\lambda}$ satisfies the equation:

$$\mathbf{0} = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{g}_i(\boldsymbol{\beta})}{1 + \boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})}. \quad (7)$$

Thus, $l_n(\boldsymbol{\beta})$ can be written:

$$l_n(\boldsymbol{\beta}) = 2 \sum_{i=1}^n \log(1 + \boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})). \quad (8)$$

In order to study the asymptotic properties of the EL statistic given by (8), let us consider following matrix: $\mathbf{A} = \mathbb{E} \left[\pi(\mathbf{X}) \dot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\beta}^0) \dot{\mathbf{f}}^t(\mathbf{X}, \boldsymbol{\beta}^0) \right]$ and under assumption (H1): $\mathbf{B} = \mathbb{E} \left[\varepsilon^2(\mathbf{X}) \pi(\mathbf{X}) \dot{\mathbf{f}}(\mathbf{X}, \boldsymbol{\beta}^0) \dot{\mathbf{f}}^t(\mathbf{X}, \boldsymbol{\beta}^0) \right]$. On suppose for the first matrix:

(HA) \mathbf{A} is a positive definite matrix.

Let us notice that \mathbf{A} is Fisher information matrix on complete data.

3 Asymptotic properties

We give first a classical result for a MAR model, lemma that turn out to be useful in the proof of the main results.

Lemma 3.1 *Under assumptions (H1), respectively (H1bis) and (H2), (H4), (HA), we have:*

$$(i) \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_i(\boldsymbol{\beta}^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \mathbf{V}),$$

(ii) $\frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\boldsymbol{\beta}^0) \mathbf{g}_i^t(\boldsymbol{\beta}^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathbf{V}$,
 (iii) $\max_{1 \leq i \leq n} \|\mathbf{g}_i(\boldsymbol{\beta}^0)\| = o_{\mathcal{P}}(n^{1/2})$,
 with $\mathbf{V} = \mathbf{B}$ for LS and $\mathbf{V} = \mathbf{A}$ for LAD method.

Let us consider following matrix:

$$\mathbf{B}_n = n^{-1} \sum_{i=1}^n \mathbf{g}_i(\boldsymbol{\beta}^0) \mathbf{g}_i^t(\boldsymbol{\beta}^0). \quad (9)$$

Following theorem gives the asymptotic distribution of the empirical log-likelihood statistic (8), evaluated at the true value.

Theorem 3.1 *Under assumptions (H1), respectively (H1bis), (H2), (H4), (HA) $l_n(\boldsymbol{\beta}^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(d)$.*

We can use Theorem 3.1 to get approximate confidence region for $\boldsymbol{\beta}$ or for testing the hypothesis: $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^0$. Since, by the proof of Theorem 3.1 (see Appendix) for the EL statistic we have:

$l_n(\boldsymbol{\beta}^0) = \mathbf{B}_n^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_i^t(\boldsymbol{\beta}^0) \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_i(\boldsymbol{\beta}^0) \right) (1 + o_{\mathcal{P}}(1))$, in order to calculate numerically $l_n(\boldsymbol{\beta}^0)$ we can use the approximation:

$$l_n^*(\boldsymbol{\beta}^0) = \mathbf{B}_n^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_i^t(\boldsymbol{\beta}^0) \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_i(\boldsymbol{\beta}^0) \right). \quad (10)$$

We state this as a corollary.

Corollary 3.1 *Under the same assumptions as in Theorem 3.1, the asymptotic distribution of $l_n^*(\boldsymbol{\beta}^0)$ is $\chi^2(d)$.*

Thus, an asymptotic $(1 - \alpha)$ confidence region for $\boldsymbol{\beta}$, based on EL statistic on complete data is: $\{\boldsymbol{\beta}; l_n^*(\boldsymbol{\beta}) \leq c_{1-\alpha;d}\}$ where $c_{1-\alpha;d}$ is the $(1 - \alpha)$ quantile of the chi-squared distribution with degrees of freedom d . It is very interesting to note that to construct the confidence region for $\boldsymbol{\beta}$ it is not necessary to calculate the Lagrange multiplier which intervenes in (8), observed data are enough.

Asymptotic normality of LS and LAD estimators calculated on complete data is given by the following result.

Theorem 3.2 (i) *Under assumption (H1), (H2), (H4), (HA) we have:*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{n,LS} - \boldsymbol{\beta}^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}).$$

(ii) *Under assumption (H1bis), (H2), (H4), (HA) we have:*

$$2\sqrt{n}e(0)\mathbf{A}^{1/2}(\hat{\boldsymbol{\beta}}_{n,LAD} - \boldsymbol{\beta}^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d).$$

These theorem allows to give the normal approximation based confidence region, expression which will be specified in Section 5. Then, on complete data, we have the choice between four statistics ($l_n^*(\beta^0)$ for LS, $l_n^*(\beta^0)$ for LAD, $\hat{\beta}_{n,LS}$, $\hat{\beta}_{n,LAD}$) to test hypotheses or to build the asymptotic confidence region of model parameter β . We see in Sections 5 and 6, by simulations and a model on real data, that approximated EL statistics are sharply superior to normal approximation given by Theorem 3.2. If error distribution presents outliers then $l_n^*(\beta)$ for LAD method is recommended, otherwise it is better to consider $l_n^*(\beta)$ for LS method. This last one will more be developed in the following section. The missing probabilities $\pi(\mathbf{X}_i)$ are estimated by a nonparametric method, this is going to allow to reconstitute the missing responses. On the observed and the reconstituted observations one defined a new EL statistic, which also satisfies a Wilk's theorem. Besides, numerically, it gives very competitive results (see Sections 5 and 6).

4 Special case of LS estimator

Following e.g. Xue (2009a), Qin et al. (2009) for linear model, we shall introduce the forecast of Y_i , constructed by using LS estimator for parameter β and a nonparametric estimator for probability $\pi(\mathbf{X}_i)$:

$$Y_{n,i} = \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)} Y_i + \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)}\right) f(\mathbf{X}_i; \hat{\beta}_{n,LS}), \quad i = 1, \dots, n, \quad (11)$$

with $\hat{\pi}(\mathbf{X}_i)$ a nonlinear estimator for $\pi(\mathbf{X}_i)$, as in the linear regression Xue (2009a):

$$\hat{\pi}(\mathbf{X}_i) = \frac{\sum_{i=1}^n \delta_i K_h((\mathbf{X}_i - \mathbf{x})/h)}{\max\{1, \sum_{i=1}^n K_h((\mathbf{X}_i - \mathbf{x})/h)\}},$$

where $h = h_n$ is a positive sequence tending to 0 as $n \rightarrow \infty$ and K_h is a kernel function defined in \mathbb{R}^d . The bandwidth h_n satisfies:

(H5) $nh_n^{2d} M_n^{-2d} \rightarrow \infty$ and $nh_n^{4\max\{2, d-1\}} \rightarrow 0$, as $n \rightarrow \infty$, with the sequence M_n given in assumption (H3).

The kernel function K , satisfies the classical condition (imposed also for the linear model using the LS method of Xue (2009a)):

(H6) there exist positive constants C_1 , C_2 and ρ such that: $C_1 \mathbb{1}_{\|u\| \leq \rho} \leq K(u) \leq C_2 \mathbb{1}_{\|u\| \leq \rho}$.

Concerning the selection probability function $\pi(\mathbf{X})$ let us make following regularity hypothesis necessary in the study of its nonparametric estimator.

(H7) $p(\mathbf{x})$ has bounded partial derivatives up to order $\max(2, d-1)$ almost everywhere.

Conditions (H5)-(H7) are usual assumptions for convergent rates of kernel estimating method.

Let us denote $\theta^0 = \mathbb{E}[Y]$ the mean of Y and $\sigma^2(\mathbf{x}) = \mathbb{E}[\varepsilon^2 | \mathbf{X} = \mathbf{x}]$ error variance.

Following lemma gives the asymptotic normality for the sequence $(Y_{n,i})$ and other two similar results of Lemma 3.1.

Lemma 4.1 *Under assumptions (H1)-(H6) and if $\text{Var}[f(\mathbf{X}, \beta^0)] < \infty$ we have:*

- (i) $\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_{n,i} - \theta^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, W),$
 - (ii) $\frac{1}{n} \sum_{i=1}^n (Y_{n,i} - \theta^0)^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} W,$
 - (iii) $\max_{1 \leq i \leq n} |Y_{n,i}| = o_{\mathbb{P}}(n^{1/2}),$
- with $W = \text{Var}[f(\mathbf{X}, \beta^0)] + \mathbb{E}\left[\frac{\sigma^2(\mathbf{X})}{\pi(\mathbf{X})}\right].$

Using similar arguments as for Theorem 3.1 we obtain the following result. We hence omit its proof.

Theorem 4.1 *Suppose that assumptions (H1)-(H6) hold, then for empirical log-likelihood for θ^0 :*

$$l_{n,Y}(\theta^0) = -2 \sup_{(p_1, \dots, p_n) \in (0,1)^n} \left\{ \sum_{i=1}^n \log(np_i); \quad \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i Y_{n,i} = \theta^0 \right\},$$

we have: $l_{n,Y}(\theta^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(1).$

This result can be used to make test of hypothesis or to construct asymptotic confidence region for the response variable. Xue (2009b) constructs a weight-corrected empirical log-likelihood ratio for θ^0 which is also asymptotically chi-squared.

Let be now following functions constructed using the reconstituted response:

$$\mathbf{g}_{n,i}(\beta) = [Y_{n,i} - f(\mathbf{X}_i; \beta)] \dot{\mathbf{f}}(\mathbf{X}_i; \beta), \quad i = 1, \dots, n.$$

Consider also the empirical log-likelihood associated at $\mathbf{g}_{n,i}(\beta)$:

$$\hat{l}_n(\beta) = -2 \sup_{(p_1, \dots, p_n) \in (0,1)^n} \left\{ \sum_{i=1}^n \log(np_i); \quad \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \mathbf{g}_{n,i}(\beta) = 0 \right\}.$$

Then, the equivalent of (8) is:

$$\hat{l}_n(\beta) = 2 \sum_{i=1}^n \log(1 + \boldsymbol{\lambda}^t \mathbf{g}_{n,i}(\beta)). \quad (12)$$

Consider following lemma needed for Theorem 4.2.

Lemma 4.2 *Under assumptions (H1)-(H7), we have:*

- (i) $n^{-1/2} \sum_{i=1}^n \mathbf{g}_{n,i}(\beta^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d \left(\mathbf{0}, \mathbb{E} \left[\pi(\mathbf{X})^{-1} \varepsilon^2(\mathbf{X}) \dot{\mathbf{f}}(\mathbf{X}; \beta^0) \dot{\mathbf{f}}^t(\mathbf{X}; \beta^0) \right] \right).$

- (ii) $n^{-1} \sum_{i=1}^n \mathbf{g}_{n,i}(\boldsymbol{\beta}^0) \mathbf{g}_{n,i}^t(\boldsymbol{\beta}^0) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbb{E} \left[\pi(\mathbf{X})^{-1} \varepsilon^2(\mathbf{X}) \dot{\mathbf{f}}(\mathbf{X}; \boldsymbol{\beta}^0) \dot{\mathbf{f}}^t(\mathbf{X}; \boldsymbol{\beta}^0) \right].$
- (iii) $\max_{1 \leq i \leq n} \|\mathbf{g}_{n,i}(\boldsymbol{\beta}^0)\| = o_{\mathbb{P}}(n^{1/2}).$

Following result shows that the empirical log-likelihood ratio on $\boldsymbol{\beta}^0$ based on the reconstituted data converges to towards $\chi^2(d)$. This theorem shows in a similar way as Theorem 3.1, then the proof will be omitted.

Theorem 4.2 *Under assumptions (H1)-(H7) we have: $\hat{l}_n(\boldsymbol{\beta}^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(d)$.*

From Theorem 4.2, one can construct an asymptotic $(1 - \alpha)$ -level confidence region for $\boldsymbol{\beta}$ using all available values for \mathbf{X} and the reconstituted values for Y .

Corollary 4.1 *In a similar way in the complete data, Corollary 3.1, the statistic $\hat{l}_n(\boldsymbol{\beta})$ may be approximated by:*

$$\hat{l}_n^*(\boldsymbol{\beta}^0) = \mathbf{B}_n^{*-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_{n,i}^t(\boldsymbol{\beta}^0) \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_{n,i}(\boldsymbol{\beta}^0) \right), \quad (13)$$

with $\mathbf{B}_n^* = n^{-1} \sum_{i=1}^n \mathbf{g}_{n,i}(\boldsymbol{\beta}^0) \mathbf{g}_{n,i}^t(\boldsymbol{\beta}^0)$. The asymptotic distribution of $\hat{l}_n^*(\boldsymbol{\beta}^0)$ is $\chi^2(d)$.

This implies that for testing hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^0$ we can use the statistic $\hat{l}_n^*(\boldsymbol{\beta}^0)$ with asymptotic reject region $\{(Y_i, \mathbf{X}_i, \delta_i)_{1 \leq i \leq n}; \hat{l}_n^*(\boldsymbol{\beta}^0) > c_{1-\alpha;d}\}$ where $c_{1-\alpha;d}$ is the $(1 - \alpha)$ quantile of the chi-squared distribution with degrees of freedom d .

Remark 1 *Since the convergence rate of $\hat{\boldsymbol{\beta}}_{n,LAD}$ to $\boldsymbol{\beta}^0$ can be slower than $n^{-1/2}$ (see Ciuperca (2010)), then Lemma 4.2 can not be true and the analogue of the Theorem 4.2 cannot be consider for LAD estimator.*

We can minimise $l_n(\boldsymbol{\beta}^0)$ and we obtain another estimator $\tilde{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}^0$, called the maximum empirical likelihood estimator (MELE). Using the same arguments as used in the proof of Theorem 1 in Qin and Lawless (1994), we obtain:

Theorem 4.3 *Under assumptions (H1), (H2), (H4), (HA) and:*

- $\mathbb{E}[\mathbf{g}_{i,LS}(\boldsymbol{\beta}^0) \mathbf{g}_{i,LS}^t(\boldsymbol{\beta}^0)]$ is positive definite,
- $\partial \mathbf{g}_{i,LS}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$ is continuous in a neighborhood of the true value $\boldsymbol{\beta}^0$, $\|\partial \mathbf{g}_{i,LS}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}\|$ and $\|\mathbf{g}_{i,LS}(\boldsymbol{\beta})\|^3$ are bounded by some integrable function in this neighborhood,
- the rank of $\mathbb{E}[\partial \mathbf{g}_{i,LS}(\boldsymbol{\beta}^0) / \partial \boldsymbol{\beta}]$ is d ,

- $\partial^2 \mathbf{g}_{i,LS}(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \boldsymbol{\beta}^t$ is continuous in a neighborhood of the true value $\boldsymbol{\beta}^0$, $\|\partial^2 \mathbf{g}_{i,LS}(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \boldsymbol{\beta}^t\|$ is bounded by some integrable function in this neighborhood

then

(i) convergence rate of $\tilde{\boldsymbol{\beta}}_n$ is $n^{-1/3}$: $\|\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^0\| = O_P(n^{-1/3})$.

(ii) $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \mathbf{V}_{MELE})$, with $\mathbf{V}_{MELE} = \mathbf{A}^{-1} \mathbf{B}_{MELE} \mathbf{A}^{-1}$, $\mathbf{B}_{MELE} = \mathbb{E}[\pi(\mathbf{X}) \varepsilon^2(\mathbf{X}) \ddot{\mathbf{f}}(\mathbf{X}; \boldsymbol{\beta}^0) \ddot{\mathbf{f}}^t(\mathbf{X}; \boldsymbol{\beta}^0)] - \mathbb{E}[\pi(\mathbf{X}) \dot{\mathbf{f}}(\mathbf{X}; \boldsymbol{\beta}^0) \dot{\mathbf{f}}^t(\mathbf{X}; \boldsymbol{\beta}^0) \dot{\mathbf{f}}(\mathbf{X}; \boldsymbol{\beta}^0) \dot{\mathbf{f}}^t(\mathbf{X}; \boldsymbol{\beta}^0)]$.

5 Simulation study

In this section we use Monte Carlo simulation to compare empirical likelihood method with normal method. For nominal confidence level $1 - \alpha = 0.95$, using the simulated samples, we evaluated the coverage probabilities (CP) of the confidence regions (CR) given by:

- approximated empirical log-likelihood method on the completely observed data (Theorem 3.1): $CR_{LS} = \{\boldsymbol{\beta}; l_n^*(\boldsymbol{\beta}) \leq c_{1-\alpha;d}\}$, $CR_{LAD} = \{\boldsymbol{\beta}; l_n^*(\boldsymbol{\beta}) \leq c_{1-\alpha;d}\}$ by LS or LAD method, respectively, where $c_{1-\alpha;d}$ is the $(1 - \alpha)$ quantile of the standard chi-square distribution with d -degrees of freedom and $l_n^*(\boldsymbol{\beta})$ given by (10);

- approximated empirical log-likelihood $\hat{l}_n^*(\boldsymbol{\beta})$, given by (13), associated at $\mathbf{g}_{n,i}(\boldsymbol{\beta})$ on the reconstituted data: $\hat{CR}_{LS} = \{\boldsymbol{\beta}; \hat{l}_n^*(\boldsymbol{\beta}) \leq c_{1-\alpha;d}\}$;

- normal method, based on Theorem 3.2:

$NCR_{LS} = \{\boldsymbol{\beta}; n(\hat{\boldsymbol{\beta}}_{n,LS} - \boldsymbol{\beta}^0)^t \mathbf{A} \mathbf{B}^{-1} \mathbf{A} (\hat{\boldsymbol{\beta}}_{n,LS} - \boldsymbol{\beta}^0) \leq c_{1-\alpha;d}\}$ and

$NCR_{LAD} = \{\boldsymbol{\beta}; 4ne(0)(\hat{\boldsymbol{\beta}}_{n,LAD} - \boldsymbol{\beta}^0)^t \mathbf{A} (\hat{\boldsymbol{\beta}}_{n,LAD} - \boldsymbol{\beta}^0) \leq c_{1-\alpha;d}\}$.

Throughout this section, the kernel function is taken as the Epanechnikov kernel $K(x) = 0.75(1 - x^2)\mathbb{1}_{|x| \leq 1}$, the bandwidth sequence $h_n = n^{-1/7}$ which satisfies assumption (H5). We generate $M = 2000$ Monte Carlo random samples of size n for $X \sim \mathcal{N}(1, 1)$ and for errors either $\mathcal{N}(0, \sigma^2)$ either Laplace $\mathcal{L}(0, \sigma^2)$ or Cauchy $\mathcal{C}(0, \sigma^2)$, with $\sigma = 1, 2$. We consider following two cases of response probability under the MAR assumption:

a) $\pi(x) = 0.8 + 0.2|x - 1|$ if $|x - 1| \leq 1$ and 0.95 elsewhere (similar the linear case of Xue (2009a)). The average missing rate is 0.91 and empirical mean of $\hat{\pi}(X_i)$ is 0.905 for $X \sim \mathcal{N}(1, 1)$.

b) $\pi(x) = 0.8$ for all x .

Case a) unlike to case b) excludes most of the error outliers.

The coverage probability are estimated by the frequency of the true values $\boldsymbol{\beta}^0$ falling into the confidence intervals in $M=2000$ simulations.

We denote by CP_{LS} , CP_{LAD} , \hat{CP}_{LS} , NCP_{LS} , NCP_{LAD} the coverage probabilities corresponding to the confidence regions CR_{LS} , CR_{LAD} , \hat{CR}_{LS} , NCR_{LS} ,

$NCRLAD$ respectively.

All simulations, calculations of estimations and statistical computations were performed using R language. To calculate the LS estimations on nonlinear model we used *nls* function of package *stats* and on linear model *lm* function of package *base*. To calculate the LAD estimations function *nls* of package *quantreg* was used if f is nonlinear and *rq* function of the same package for linear model. We used also *VGAM* package for random generation of the Laplace distribution by *rlaplace* function. To generate Normal and Cauchy distribution the functions *rmnorm* respectively *rcauchy* of *stats* package are used.

5.1 Nonlinear model

Let us consider following nonlinear function corresponding a two-compartment model:

$$f(x; \boldsymbol{\beta}) = \frac{\beta_1}{\beta_1 - \beta_2} (e^{-\beta_2 x} - e^{-\beta_1 x}), \quad \boldsymbol{\beta} = (\beta_1, \beta_2),$$

with the true values: $\beta_1^0 = 1$ and $\beta_2^0 = 1.5$.

In Figures 1, 2, 3 a simulation of this model is plotted for $n = 300$, $\pi(x) = 0.8$, $\sigma = 1$ for errors $\varepsilon \sim \mathcal{N}, \mathcal{L}, \mathcal{C}$ respectively. We represented with “solid circle” the reconstituted values $Y_{n,i}$ and with “triangle” the true complete values Y_i . In Table 1 we get the coverage probability CP_{LS} , CP_{LAD} , \hat{CP}_{LS} , NCP_{LS} , NCP_{LAD} in the case $\pi(X) = 0.8 + 0.2|x - 1|$ if $|x - 1| \leq 1$ and 0.95 elsewhere when $n = 300$, $n = 100$. In Table 2 we get the same five coverage probabilities for the case $\pi(X) = 0.8$. These two tables show that:

- (1) CP_{LS} is better than CP_{LAD} excepting for the case $\varepsilon \sim \mathcal{C}(0, 2)$;
- (2) coverage probabilities on the reconstituted responses \hat{CP}_{LS} is bigger than 0.95 and has values very closed to values of CP_{LS} obtained on complete data;
- (3) NCP_{LS} , NCP_{LAD} are worse than coverage rate on the complete or reconstituted data, particularly for $n = 100$ or Cauchy errors. For Cauchy distribution we cannot calculate NCP_{LS} since the assumptions of Theorem 3.2(i), more precisely (H1), are not satisfied ;
- (4) Since case a) has tendency to eliminate the outliers of ε , the rates of normal coverage probability are bigger for a) than for b).

Tables 3 and 4 contain empirical mean, standard-deviation of parameter estimations $\hat{\boldsymbol{\beta}}_{n,LS}$ and $\hat{\boldsymbol{\beta}}_{n,LAD}$ for the two cases of response probability. We deduce from these two tables:

- (1) there is no difference between the parameter estimations obtained for both cases of probability a) and b), for the same n and the same distri-

- bution of ε ;
- (2) already known thing: for distributions with outliers (Laplace or Cauchy here) the LAD estimators have a standard-deviation smaller than LS estimators.

Table 5: whether it is for the probability a) or b), the medians of LAD estimations are more close to the true values than the medians of LS estimations, if ε have outliers. The precision of the estimators increases with n and in a less measure with the probability $\pi(x)$.

5.2 Linear model

Consider now the simple linear model: $Y_i = X_i\beta + \varepsilon_i$, with the true value of the parameter $\beta^0 = 10$. In Tables 6 and 7 we get the coverage probabilities for the two case of $\pi(X)$ and for $n = 100, 20$. From these two tables we make the following remarks:

- (1) in general CP_{LAD} performs better than CP_{LS} when $\varepsilon \sim \mathcal{L}$ or \mathcal{C} ;
- (2) \hat{CP}_{LS} is bigger than 0.95 and has very close values to CP_{LS} values;
- (3) similar of (3) for nonlinear case, except that $NCP_{LAD} < NCP_{LS}$;
- (4) with regard to both cases of probability $\pi(X)$: there is no difference between the three coverage probabilities by estimated EL methods, contrary to the nonlinear case nor between the natural coverage probabilities.

Tables 8 and 9: the standard-deviation of LS estimators is much bigger than for LAD's when $\varepsilon \sim \mathcal{C}$. That comes from the outliers presence in nonlinear case, the algorithm to find the LS estimator can not converge, thus the estimations are not available.

5.3 Conclusion

The coverage probabilities of CR given by normal method are lower than the nominal level $1 - \alpha$ especially for small sample sizes. The simulation results show that in terms of coverage probability the empirical likelihood methods outperform the normal approximation method in particular when error distribution have outliers or its has a bigger standard-deviation.

It is interesting to note that, if the distribution of the errors presents outliers the approximated EL for LAD method gives generally best results than that for LS method that concerns EL coverage probability or parameter estimators efficiency. Then, if $\varepsilon \sim \mathcal{N}$, in order to test hypothesis $H_0 : \beta = \beta^0$, it is recommended to take the test statistic $l_n^*(\beta)$ given by (10) for complete data or $\hat{l}_n^*(\beta)$, given by (13) for improved data using LS method. Let us emphasize

that $\hat{l}_n^*(\boldsymbol{\beta})$ gives very close results to those of $l_n^*(\boldsymbol{\beta})$. On the other hand, if ε presents outliers, to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^0$ it is recommended $l_n^*(\boldsymbol{\beta})$ given by (10) at $\mathbf{g}_i = \mathbf{g}_{i,LAD}$.

6 An application

In R language consider *Chwirut1* data of *NISTnls* package to model the ultrasonic response value Y function to metal distance value X using nonlinear function:

$$f(X, \boldsymbol{\beta}) = \frac{\exp(-b_1 X)}{b_2 + b_3 X}, \quad \boldsymbol{\beta} = (b_1, b_2, b_3).$$

The realizations of (Y, X) are known for $n = 214$ observations. In Figure 4 we represented variable Y =ultrasonic response function of regressor X =metal distance: “solid circle” for reconstituted values $Y_{n,i}$, “triangle” for true complete values Y_i . The LS parameter estimations on all 214 observations are: $\hat{b}_1 = 0.19$, $\hat{b}_2 = 0.006$, $\hat{b}_3 = 0.10$. By 10000 Monte Carlo samples, we eliminate 20%, 50%, 80% values of Y . Consider that the true value $\boldsymbol{\beta}^0$ of $\boldsymbol{\beta}$ is the one obtained on 214 observations. In Table 10 we have the acceptance rate of hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^0$ with respect to missing probability $1 - \pi(X)$ and with respect to the estimation method (LS or LAD) on complete data using statistic $l_n^*(\boldsymbol{\beta})$ given by (10), on reconstituted data using statistic $\hat{l}_n^*(\boldsymbol{\beta})$ given by (13). We observe that, even by eliminating 80% observations, the obtained estimations are very close to $\boldsymbol{\beta}^0$. In Table 11 we find the empirical means, standard-deviations of differences $Y_i - \hat{Y}_{i,0}$, $Y_i - \hat{Y}_i$, where: Y_i are the true values of Y , $\hat{Y}_{i,0}$ is the forecast for Y_i using all 214 observations by LS method, $\hat{Y}_{i,0} = f(X_i, \hat{\boldsymbol{\beta}}_{LS})$. For $i = 1, \dots, n$, $\hat{Y}_i = Y_{ni}$ is the reconstituted value of Y_i by relation (11). We deduce that the variable response is better reconstituted by (11) than by LS method on all observations.

7 Appendix: proofs

We give proofs for the results in Sections 3 and 4.

In the following, we denote by C a generic positive finite constant not depending on n which may take different values in different formulae or even in different parts of the same formula.

Proof of Lemma 3.1 (i) We apply the central limit theorem.

(ii) By the weak law of large numbers.

(iii) We combine Lemma 3 of Owen (1990) and results for nonlinear regression without missing data (see e.g. Seber and Wild (2003)). ■

Proof of Theorem 3.1 We take the Taylor's expansion of $l_n(\beta^0)$ given by (8):
 $l_n(\beta^0) = 2\lambda^t \sum_{i=1}^n \mathbf{g}_i(\beta^0) - \lambda^t \sum_{i=1}^n \mathbf{g}_i(\beta^0) \mathbf{g}_i^t(\beta^0) \lambda + \lambda^t O_P(\sum_{i=1}^n \mathbf{g}_i(\beta^0) (\lambda^t \mathbf{g}_i(\beta^0))^2)$.
 By a similar approach of Owen (1990), using also Lemma 3.1, we obtain:
 $\|\lambda\| = O_P(n^{-1/2})$. Using Lemma 3.1(iii) we obtain:

$$n^{-1} \lambda^t \sum_{i=1}^n \mathbf{g}_i(\beta^0) (\lambda^t \mathbf{g}_i(\beta^0))^2 = o_P \left(n^{-1} \sum_{i=1}^n (\lambda^t \mathbf{g}_i(\beta^0))^2 \right). \quad (14)$$

Then, returning at $l_n(\beta^0)$, we have:

$$l_n(\beta^0) = 2 \sum_{i=1}^n \left[\lambda^t \mathbf{g}_i(\beta^0) - \frac{1}{2} \lambda^t \mathbf{g}_i(\beta^0) \mathbf{g}_i^t(\beta^0) \lambda \right] (1 + o_P(1)). \quad (15)$$

Relation (7) can be written also:

$$\mathbf{0} = \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\beta^0) - \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\beta^0) \mathbf{g}_i^t(\beta^0) \lambda + \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\beta^0) (\lambda^t \mathbf{g}_i(\beta^0))^2 (1 + \lambda^t \mathbf{g}_i(\beta^0))^{-1}.$$

On the other hand, using (14), we obtain:

$$\mathbf{0} = \left[\frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\beta^0) - \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\beta^0) \mathbf{g}_i^t(\beta^0) \lambda \right] (1 + o_P(1)). \quad (16)$$

Thus

$$\lambda = \left[\frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\beta^0) \mathbf{g}_i^t(\beta^0) \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{g}_i(\beta^0) \right] (1 + o_P(1)). \quad (17)$$

Relation (16) implies also: $\lambda^t \sum_{i=1}^n \mathbf{g}_i(\beta^0) = \left[\lambda^t \sum_{i=1}^n \mathbf{g}_i(\beta^0) \mathbf{g}_i^t(\beta^0) \lambda \right] (1 + o_P(1))$. Then, expression (15) of $l_n(\beta^0)$ becomes, using relation (17):

$$l_n(\beta^0) = \left[\lambda^t \sum_{i=1}^n \mathbf{g}_i(\beta^0) \right] (1 + o_P(1)) = \mathbf{B}_n^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_i^t(\beta^0) \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{g}_i(\beta^0) \right) (1 + o_P(1)).$$

By Lemma 3.1(i) and (ii) we have: $l_n(\beta^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(d)$. ■

Proof of Theorem 3.2.(i) By definition, the LS estimator is obtained as the solution of the system: $\mathbf{0} = \sum_{i=1}^n \mathbf{g}_i(\hat{\beta}_{n,LS})$. Using Lemma 3.1(i) we have:
 $n^{-1/2} \sum_{i=1}^n \mathbf{g}_i(\beta^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \mathbf{B})$, then $n^{-1/2} \sum_{i=1}^n [\mathbf{g}_i(\hat{\beta}_{n,LS}) - \mathbf{g}_i(\beta^0)] \xrightarrow[n \rightarrow \infty]{\mathcal{L}}$

$\mathcal{N}(0, \mathbf{B})$. This last relation implies:

$$n^{-1} \sum_{i=1}^n [\mathbf{g}_i(\hat{\beta}_{n,LS}) - \mathbf{g}_i(\beta^0)] = O_P(n^{-1/2}). \quad (18)$$

On the other hand, using Taylor expansion:

$$\begin{aligned} \mathbf{g}_i(\hat{\beta}_{n,LS}) - \mathbf{g}_i(\beta^0) &= \delta_i \dot{\mathbf{f}}(\mathbf{X}_i; \hat{\beta}_{n,LS}) [Y_i - f(\mathbf{X}_i; \hat{\beta}_{n,LS})] - \delta_i \dot{\mathbf{f}}(\mathbf{X}_i; \beta^0) [Y_i - f(\mathbf{X}_i; \beta^0)] \\ &= \delta_i \varepsilon_i [\dot{\mathbf{f}}(\mathbf{X}_i; \hat{\beta}_{n,LS}) - \dot{\mathbf{f}}(\mathbf{X}_i; \beta^0)] + \delta_i [\dot{\mathbf{f}}(\mathbf{X}_i; \hat{\beta}_{n,LS}) - \dot{\mathbf{f}}(\mathbf{X}_i; \beta^0)] [f(\mathbf{X}_i; \beta^0) - f(\mathbf{X}_i; \hat{\beta}_{n,LS})] \\ &\quad + \delta_i \dot{\mathbf{f}}(\mathbf{X}_i; \beta^0) [f(\mathbf{X}_i; \beta^0) - f(\mathbf{X}_i; \hat{\beta}_{n,LS})] \end{aligned}$$

Since $\hat{\beta}_{n,LS} \xrightarrow[n \rightarrow \infty]{a.s.} \beta^0$, we obtain: $\mathbf{g}_i(\hat{\beta}_{n,LS}) - \mathbf{g}_i(\beta^0) = \delta_i \varepsilon_i \ddot{\mathbf{f}}(\mathbf{X}_i; \beta_{1,n})(\hat{\beta}_{n,LS} - \beta^0) + (\hat{\beta}_{n,LS} - \beta^0) \delta_i \dot{\mathbf{f}}^t(\mathbf{X}_i; \beta_{2,n}) \ddot{\mathbf{f}}(\mathbf{X}_i; \beta_{1,n})(\hat{\beta}_{n,LS} - \beta^0) + (\hat{\beta}_{n,LS} - \beta^0) \delta_i \dot{\mathbf{f}}^t(\mathbf{X}_i; \beta_{2,n}) \dot{\mathbf{f}}(\mathbf{X}_i; \beta^0)$, with $\beta_{1,n} = \beta^0 + u_1(\hat{\beta}_{n,LS} - \beta^0)$, $\beta_{2,n} = \beta^0 + u_2(\hat{\beta}_{n,LS} - \beta^0)$, $u_1, u_2 \in [0, 1]$.

Under assumptions (H1) and (H4):

$$n^{-1} \sum_{i=1}^n [\mathbf{g}_i(\hat{\beta}_{n,LS}) - \mathbf{g}_i(\beta^0)] = [o_P(\hat{\beta}_{n,LS} - \beta^0) + (\hat{\beta}_{n,LS} - \beta^0) A] (1 + o_P(1)).$$

Taking into account relation (18), we have:

$$\hat{\beta}_{n,LS} - \beta^0 = O_P(n^{-1/2}). \quad (19)$$

Combining these last two results, we obtain:

$$\sqrt{n}(\hat{\beta}_{n,LS} - \beta^0) A (1 + o_P(1)) = n^{-1/2} \sum_{i=1}^n [\mathbf{g}_i(\hat{\beta}_{n,LS} - \beta^0)] \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \mathbf{B})$$

and claim follows.

(ii) We apply Weiss (1991). ■

Proof of Lemma 4.1 (i) Let us consider following decomposition:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_{n,i} - \theta^0) \equiv S_1 + S_2 + S_3, \text{ with:}$$

$$S_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\delta_i \varepsilon_i}{\pi(\mathbf{X}_i)} + f(\mathbf{X}_i; \beta^0) - \theta^0 \right],$$

$$S_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \varepsilon_i \left[\frac{1}{\hat{\pi}(\mathbf{X}_i)} - \frac{1}{\pi(\mathbf{X}_i)} \right],$$

$$S_3 = \frac{1}{\sqrt{n}} \sum_{i=1}^n [f(\mathbf{X}_i; \hat{\beta}_{n,LS}) - f(\mathbf{X}_i; \beta^0)] \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)} \right).$$

$$\text{For } S_1: \text{Var} \left[\frac{\delta_i \varepsilon_i}{\pi(\mathbf{X}_i)} + f(\mathbf{X}_i; \beta^0) - \theta^0 \right] = \text{Var} \left[\frac{\delta_i \varepsilon_i}{\pi(\mathbf{X}_i)} \right] + \text{Var} [f(\mathbf{X}_i; \beta^0)] + 2 \text{Cov} (f(\mathbf{X}_i; \beta^0), \frac{\delta_i \varepsilon_i}{\pi(\mathbf{X}_i)}).$$

By (H1) we have $\text{Cov} (f(\mathbf{X}_i; \beta^0), \frac{\delta_i \varepsilon_i}{\pi(\mathbf{X}_i)}) = 0$. On the other hand: $\text{Var} \left[\frac{\delta_i \varepsilon_i}{\pi(\mathbf{X}_i)} \right] =$

$$\mathbb{E} \left[\frac{\delta_i^2 \varepsilon_i^2}{\pi^2(\mathbf{X}_i)} \right] = \mathbb{E} \left[\frac{\delta_i \varepsilon_i^2}{\pi^2(\mathbf{X}_i)} \right] = \mathbb{E} \left[\frac{\varepsilon_i^2(\mathbf{X})}{\pi(\mathbf{X})} \right].$$

Since $(\mathbf{X}_i, \varepsilon_i)$ are independent for different i , then the random variables $\frac{\delta_i \varepsilon_i}{\pi(\mathbf{X}_i)} - f(\mathbf{X}_i; \beta^0) - \theta^0$ are also independent.

Then, we can apply the central limit theorem: $S_1 \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, W)$.

For S_3 . We make a limited development for S_3 until order 2 around β^0 and taking into account relation (19), hypothesis $\pi(\mathbf{x}) > 0$ and (H4):

$$S_3 = (\hat{\beta}_{n,LS} - \beta^0)^t \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\pi(\mathbf{X}_i)} \right) \dot{\mathbf{f}}(\mathbf{X}_i; \beta^0) (1 + o_P(1)) = O_P(n^{-1/2}) o_P(n^{1/2}) = o_P(1).$$

By Lemma 3 of Xue (2009a) we have $S_2 = o_P(1)$, then the claim (i) is

proved. The proof of (ii) and (iii) is similar. \blacksquare

Proof of Lemma 4.2 (i) Function $g_{n,i}(\boldsymbol{\beta})$ can be written:

$$\mathbf{g}_{n,i}(\boldsymbol{\beta}) = \frac{\mathbf{g}_{i,LS}(\boldsymbol{\beta})}{\hat{\pi}(\mathbf{X}_i)} + \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)}\right) [f(\mathbf{X}_i; \hat{\boldsymbol{\beta}}_{n,LS}) - f(\mathbf{X}_i; \boldsymbol{\beta})] \dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}). \quad (20)$$

For the first term of the right-hand side of (20) we have:

$$n^{-1/2} \sum_{i=1}^n \frac{\mathbf{g}_{i,LS}(\boldsymbol{\beta}^0)}{\hat{\pi}(\mathbf{X}_i)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_d \left(\mathbf{0}, \mathbb{E} \left[\pi(\mathbf{X})^{-1} \varepsilon^2(\mathbf{X}) \dot{\mathbf{f}}(\mathbf{X}; \boldsymbol{\beta}^0) \dot{\mathbf{f}}^t(\mathbf{X}; \boldsymbol{\beta}^0) \right] \right). \quad (21)$$

On the other hand, using a limited development, relation (19) and assumption (H4): $n^{-1/2} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)}\right) [f(\mathbf{X}_i; \hat{\boldsymbol{\beta}}_{n,LS}) - f(\mathbf{X}_i; \boldsymbol{\beta}^0)] \dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)$
 $= \frac{\hat{\boldsymbol{\beta}}_{n,LS} - \boldsymbol{\beta}^0}{\sqrt{n}} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)}\right) \dot{\mathbf{f}}^t(\mathbf{X}_i; \boldsymbol{\beta}_{1,n}) \dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)$, with $\boldsymbol{\beta}_{1,n}$ between $\boldsymbol{\beta}^0$ and $\hat{\boldsymbol{\beta}}_{n,LS}$. Let us consider following random variable:

$S_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)}\right) \dot{\mathbf{f}}^t(\mathbf{X}_i; \boldsymbol{\beta}) \dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)$, for $\boldsymbol{\beta}$ in a $n^{-1/2}$ -neighborhood of $\boldsymbol{\beta}^0$. If we prove:

$$S_n(\boldsymbol{\beta}) = o_P(1), \quad (22)$$

with $o_P(1)$ uniformly in $\boldsymbol{\beta}$, tacking into account also (19), we obtain (i).

Let us remind at first some results of the paper Xue (2009a), on the estimators $\hat{\pi}(\mathbf{X}_i)$ of $\pi(\mathbf{X}_i)$. Under assumptions (H3), (H6) and (H7) we have uniformly over $1 \leq i \leq n$:

$$\mathbb{E}[\hat{\pi}(\mathbf{X}_i) - \pi(\mathbf{X}_i)]^2 = O((nh_n^d)^{-1} M_n^d) + O(h_n^{2\max(2,d-1)}) + o(n^{-1/2}). \quad (23)$$

In the same paper, following two random variables are considered: $W_{nj}(\mathbf{x}) = K_h(\mathbf{X}_k - \mathbf{x}) / \max\{1, \sum_{i=1}^n K_h(\mathbf{X}_i - \mathbf{x})\}$, for $\mathbf{x} \in \Upsilon$, $j = 1, \dots, n$ and $C_n(\mathbf{X}_i) = \max\{1, C_1 \sum_{j \neq i} \mathbb{1}_{\|\mathbf{X}_j - \mathbf{X}_i\| \leq \rho h_n}\}$ with C_1, ρ positive constants defined in assumption (H6). For this last random variable we have $\mathbb{E}[C_n^{-1}(\mathbf{X}_i)] = o_P(n^{-1/2})$.

We now turn to study $S_n(\boldsymbol{\beta})$, which can be written, after a limited development:

$$S_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)}\right) \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 + \frac{\boldsymbol{\beta} - \boldsymbol{\beta}^0}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)}\right) \ddot{\mathbf{f}}(\mathbf{X}_i; \tilde{\boldsymbol{\beta}}) \dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0) \quad (24)$$

with $\tilde{\boldsymbol{\beta}}$ between $\boldsymbol{\beta}^0$ and $\boldsymbol{\beta}$. We denote decomposition (24): $S_n(\boldsymbol{\beta}) \equiv S_1 + S_2$. Using Cauchy-Schwarz inequality, relation (19) and assumption (H4), it is easily shown that $S_2 = o_P(1)$ uniformly in $\boldsymbol{\beta}$. We study now $S_1 = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{X}_i)}\right) \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 + \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\pi(\mathbf{X}_i)} - \frac{1}{\hat{\pi}(\mathbf{X}_i)}\right) \|\ddot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 \equiv S_{11} +$

S_{12} . Second term S_{12} can be written: $S_{12} = n^{-1} \sum_{i=1}^n \delta_i \frac{\hat{\pi}(\mathbf{X}_i) - \pi(\mathbf{X}_i)}{\pi^2(\mathbf{X}_i)} \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 - n^{-1} \sum_{i=1}^n \frac{(\hat{\pi}(\mathbf{X}_i) - \pi(\mathbf{X}_i))^2}{\pi^2(\mathbf{X}_i) \hat{\pi}(\mathbf{X}_i)} \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2$. The last term of S_{12} is $o_P(1)$ by (23). For the first term of S_{12} we have the decomposition:
 $n^{-1} \sum_{i=1}^n \delta_i \frac{\hat{\pi}(\mathbf{X}_i) - \pi(\mathbf{X}_i)}{\pi^2(\mathbf{X}_i)} \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 \equiv T_1 + T_2 + T_3$, with
 $T_1 = n^{-1} \sum_{i=1}^n \frac{\delta_i}{\pi^2(\mathbf{X}_i)} \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 \sum_{j=1}^n W_{nj}(\mathbf{X}_i) [\pi(\mathbf{X}_i) - \pi(\mathbf{X}_j)]$,
 $T_2 = n^{-1} \sum_{i=1}^n \frac{\delta_i}{\pi^2(\mathbf{X}_i)} \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 \sum_{j=1}^n W_{nj}(\mathbf{X}_i) [\pi(\mathbf{X}_j) - \delta_j]$,
 $T_3 = n^{-1} \sum_{i=1}^n \frac{\delta_i}{\pi(\mathbf{X}_i)} \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 \left[1 - \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\right]$.
By Cauchy-Schwarz inequality, assumptions (H4), (H7) we have:
 $\mathbb{E}[T_1^2] \leq \frac{C}{n^2} \sum_{i=1}^n \mathbb{E} \left[\mathbb{E} \left[\sum_{j=1}^n W_{nj}^2(\mathbf{X}_i) [\pi(\mathbf{X}_i) - \pi(\mathbf{X}_j)]^2 | \mathbf{X}_i \right] \right]$
 $\leq \frac{C}{n^2} \sum_{i=1}^n \mathbb{E} \left[\mathbb{E} \left[\sum_{j=1}^n W_{nj}^2(\mathbf{X}_i) [\mathbf{X}_i - \mathbf{X}_j]^2 | \mathbf{X}_i \right] \right] \leq Ch_n^2 n^{-1} \rightarrow 0$ for $n \rightarrow \infty$.
Then $T_1 = o_P(1)$.
For T_2 : $\mathbb{E}[T_2] \leq \frac{C}{n^2} \sum_{i=1}^n \mathbb{E} \left[\sum_{j=1}^n W_{nj}^2(\mathbf{X}_i) \right] \leq \frac{C}{n^2} \sum_{i=1}^n \mathbb{E} \left[\frac{1}{C_n(\mathbf{X}_i)} \right] \rightarrow 0$ for $n \rightarrow \infty$, then $T_2 = o_P(1)$. In a similar way:
 $\mathbb{E}[T_3] \leq \frac{C}{n^2} \sum_{i=1}^n \left[\mathbb{E} \left(\frac{\delta_i}{\pi^2(\mathbf{X}_i)} \|\dot{\mathbf{f}}(\mathbf{X}_i; \boldsymbol{\beta}^0)\|^2 | \mathbf{X}_i \right) \left(1 - \sum_{j=1}^n W_{nj}(\mathbf{X}_i) \right) \right] \rightarrow 0$. Thus
 $T_1 + T_2 + T_3 = o_P(1)$, what implies $S_{12} = o_P(1)$.
Finally, for S_{11} : $\mathbb{E}[S_{11}] = 0$ and with the same arguments as for T_2 : $\mathbb{E}[S_{11}^2] \rightarrow 0$ for $n \rightarrow n$, then $S_{11} = o_P(1)$. With all this, relation (22) is proved.
the proof of (ii) and (iii) is similar. ■

References

- Bai, J., 1998, Estimation of multiple-regime regressions with least absolute deviation. *Journal of Statistical Planning Inference*, **74**, pp. 103-134.
- Ciuperca, G., 2010., Estimating nonlinear regression with and without change-points by the LAD-method. *Annals of the Institute of Statistical Mathematics*, DOI: 10.1007/s10463-009-0256-y.
- Kim, H.K., Choi, S.H., 1995, Asymptotic properties of non-linear least absolute deviation estimators. *Journal of the Korean Statistical Society*, **24**, pp. 127-139.
- Liang H., Qin Y., Zhang X., Ruppert D., 2009, Empirical likelihood-based inferences for generalized partially linear models. *Scandinavian Journal of Statistics*, **36**(3), 433-443.
- Oberhofer, W., 1982, The consistency of nonlinear regression minimizing the L_1 -norm. *The Annals of Statistics*, **10**, No. 1, pp. 316-319.
- Owen A., 1990, Empirical likelihood ratio confidence regions, *Annals of Statistics* **18**(1), 90-120.
- Qin Y., Li L., Lei Q., 2009, Empirical likelihood for linear regression models with missing responses, *Statistics and Probability Letters* **79**(11), 1391-1396.
- Qin J., Lawless J., 1994, Empirical likelihood and general estimating equations, *Annals of Statistics* **22**(1), 300-325.

- Seber G.A.F., Wild C.J., 2003, Nonlinear regression, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., Hoboken, New Jersey.
- Sun Z., Wang, Q., Dai P., 2009, Model checking for partially linear models with missing responses at random. *Journal of Multivariate Analysis*, **100**(4), 636-651.
- Sun Z., Wang, Q., 2009, Checking the adequacy of a general linear model with responses missing at random. *Journal of Statistical Planning Inference*, **139**(10), 3588-3604.
- Wang, Q., Rao J.N.K., 2002, Empirical likelihood-based inference in linear models with missing data. *Scandinavian Journal of Statistics*, **29**(3), 563-576.
- Wang, Q., Linton O., Härdle W., 2004, Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association*, **99**(466), 334-345.
- Wang, Q., Sun Z., 2007, Estimation in partially linear models with missing responses at random. *Journal of Multivariate Analysis*, **98**(7), 1470-1493.
- Weiss, A.A., 1991, Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory*, **7**, 46-68.
- Xue L., 2009, Empirical likelihood for linear models with missing responses, *Journal of Multivariate Analysis* **100**, 1353-1366.
- Xue L., 2009, Empirical likelihood confidence intervals for response mean with data missing at random, *Scandinavian Journal of Statistics* **36**(4), 671-685.
- Yang Y., Xue L., Cheng W., 2009, Empirical likelihood for a partially linear model with covariate data missing at random. *Journal of Statistical Planning Inference*, **139**(12), 4143-4153.
- Zhao Y., Chen F. 2008, Empirical likelihood inference for censored median regression model via nonparametric kernel estimation, *Journal of Multivariate Analysis* **99**(2), 215-231.

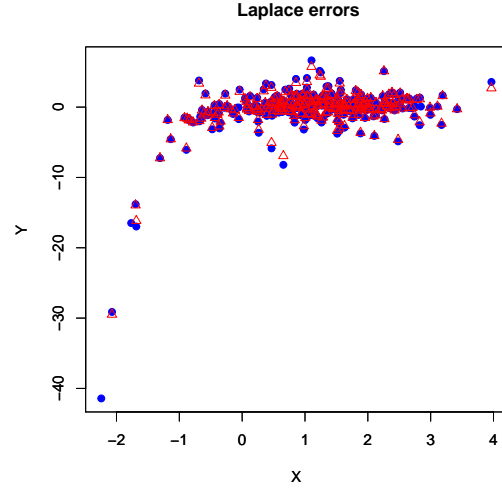
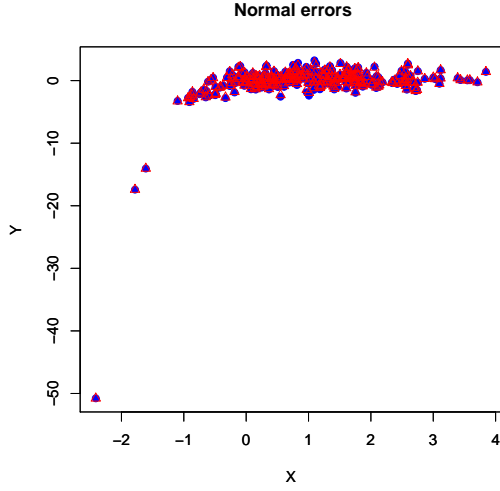


Fig. 1. Response variable Y function regressor X : “solide circle” for reconstituted values $Y_{n,i}$, “triangle” for true complete values Y_i . Normal error $\varepsilon \sim \mathcal{N}(0, 1)$, $n=300$.

Fig. 2. Response variable Y function regressor X : “solide circle” for reconstituted values $Y_{n,i}$, “triangle” for true complete values Y_i . Laplace error $\varepsilon \sim \mathcal{L}(0, 1)$, $n=300$.

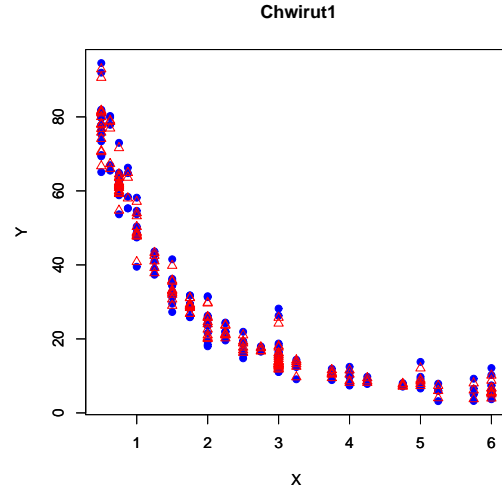
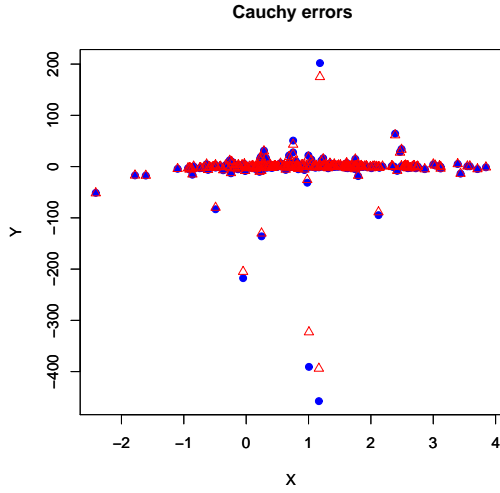


Fig. 3. Response variable Y function regressor X : “solide circle” for reconstituted values $Y_{n,i}$, “triangle” for true complete values Y_i . Cauchy error $\varepsilon \sim \mathcal{C}(0, 1)$, $n=300$.

Fig. 4. Variable Y =ultrasonic response function regressor X =metal distance: “solide circle” for reconstituted values $Y_{n,i}$, “triangle” for true complete values Y_i .

Table 1

Coverage probabilities for nonlinear model, $\pi(X) = 0.8 + 0.2|x - 1|$, if $|x - 1| \leq 1$ and 0.95 elsewhere, $n = 300, 100$, error Gaussian, Laplace, Cauchy.

ε	$n = 300$					$n = 100$				
	CP_{LS}	CP_{LAD}	\hat{CP}_{LS}	NCP_{LS}	NCP_{LAD}	CP_{LS}	CP_{LAD}	\hat{CP}_{LS}	NCP_{LS}	NCP_{LAD}
$\mathcal{N}(0, 1)$	0.977	0.825	0.959	0.981	0.884	0.975	0.747	0.956	0.63	0.507
$\mathcal{L}(0, 1)$	0.971	0.881	0.967	0.991	0.829	0.978	0.836	0.95	0.604	0.453
$\mathcal{C}(0, 1)$	0.983	0.991	0.984		0.293	0.987	0.963	0.987		0.586
$\mathcal{N}(0, 2)$	0.964	0.964	0.945	0.774	0.423	0.976	0.957	0.917	0.86	0.686
$\mathcal{L}(0, 2)$	0.977	0.979	0.974	0.455	0.394	0.978	0.959	0.978	0.925	0.439
$\mathcal{C}(0, 2)$	0.986	0.999	0.988		0.615	0.987	0.994	0.987		0.541

Table 2

Coverage probabilities for nonlinear model, $\pi(X) = 0.8$, $n = 300, 100$, error Gaussian, Laplace, Cauchy.

ε	$n = 300$					$n = 100$				
	CP_{LS}	CP_{LAD}	\hat{CP}_{LS}	NCP_{LS}	NCP_{LAD}	CP_{LS}	CP_{LAD}	\hat{CP}_{LS}	NCP_{LS}	NCP_{LAD}
$\mathcal{N}(0, 1)$	0.970	0.807	0.960	0.955	0.923	0.972	0.780	0.931	0.508	0.393
$\mathcal{L}(0, 1)$	0.972	0.872	0.975	0.986	0.891	0.976	0.823	0.975	0.227	0.392
$\mathcal{C}(0, 1)$	0.985	0.985	0.987		0.753	0.987	0.958	0.987		0.772
$\mathcal{N}(0, 2)$	0.979	0.968	0.949	0.560	0.298	0.979	0.958	0.886	0.159	0.254
$\mathcal{L}(0, 2)$	0.979	0.977	0.969	0.668	0.288	0.979	0.965	0.973	0.353	0.234
$\mathcal{C}(0, 2)$	0.986	0.999	0.985		0.636	0.991	0.992	0.990		0.699

Table 3

Mean, standard deviation of parameter estimations for nonlinear model, $\pi(X) = 0.8 + 0.2|x - 1|$ if $|x - 1| \leq 1$ and 0.95 elsewhere, error Gaussian, Laplace, Cauchy. True values $\beta_1^0 = 1$, $\beta_2^0 = 1.5$.

n	β_1				β_2			
	mean	mean	sd	sd	mean	mean	sd	sd
error	LS	LAD	LS	LAD	LS	LAD	LS	LAD
$n = 300$								
$\mathcal{N}(0, 1)$	1	1	0.13	0.16	1.50	1.49	0.26	0.31
$\mathcal{L}(0, 1)$	1	1	0.17	0.14	1.48	1.50	0.36	0.28
$\mathcal{C}(0, 1)$	0.99	0.99	0.68	0.15	1.40	1.50	1.58	0.33
$\mathcal{N}(0, 2)$	1	0.99	0.25	0.28	1.49	1.51	0.53	0.62
$\mathcal{L}(0, 2)$	0.99	0.98	0.31	0.26	1.51	1.52	0.74	0.61
$\mathcal{C}(0, 2)$	1.01	0.95	0.82	0.36	1.43	1.61	2.06	1.02
$n = 100$								
$\mathcal{N}(0, 1)$	0.97	0.96	0.25	0.29	1.58	1.59	0.70	0.77
$\mathcal{L}(0, 1)$	0.96	0.95	0.33	0.29	1.60	1.61	0.97	0.86
$\mathcal{C}(0, 1)$	1.01	0.94	0.79	0.83	1.54	1.66	2.32	2.09
$\mathcal{N}(0, 2)$	0.96	0.94	0.42	0.54	1.66	1.70	1.40	1.71
$\mathcal{L}(0, 2)$	0.93	0.91	0.48	0.65	1.74	1.81	1.65	1.93
$\mathcal{C}(0, 2)$	1.02	0.87	0.93	0.90	1.53	1.85	2.53	2.21

Table 4

Mean, standard deviation of parameter estimations for nonlinear model, $\pi(X) = 0.8$, error Gaussian, Laplace, Cauchy. True values $\beta_1^0 = 1$, $\beta_2^0 = 1.5$.

n error	β_1		sd		β_2		sd	
	mean LS	mean LAD	LS	LAD	mean LS	mean LAD	LS	LAD
$n = 300$								
$\mathcal{N}(0, 1)$	1	0.99	0.14	0.18	1.50	1.50	0.29	0.35
$\mathcal{L}(0, 1)$	1	0.99	0.19	0.15	1.50	1.50	0.40	0.31
$\mathcal{C}(0, 1)$	0.99	0.96	0.70	0.25	1.51	1.58	1.87	1.69
$\mathcal{N}(0, 2)$	1	0.98	0.26	0.31	1.49	1.52	0.58	0.70
$\mathcal{L}(0, 2)$	0.97	0.96	0.34	0.29	1.54	1.57	0.82	0.72
$\mathcal{C}(0, 2)$	0.96	0.96	0.83	0.37	1.57	1.55	2.13	0.95
$n = 100$								
$\mathcal{N}(0, 1)$	0.96	0.95	0.28	0.35	1.62	1.66	0.99	1.21
$\mathcal{L}(0, 1)$	0.96	0.95	0.32	0.31	1.62	1.64	1.23	0.94
$\mathcal{C}(0, 1)$	0.97	0.95	0.81	0.99	1.59	1.73	2.32	1.98
$\mathcal{N}(0, 2)$	0.94	0.94	0.44	0.52	1.70	1.77	1.51	1.85
$\mathcal{L}(0, 2)$	0.95	0.93	0.51	0.51	1.66	1.75	1.66	1.65
$\mathcal{C}(0, 2)$	1.09	0.89	0.94	0.58	1.25	1.83	2.28	2.04

Table 5

Median of parameter estimations for nonlinear model, two cases for $\pi(X)$, error Gaussian, Laplace or Cauchy. True values $\beta_1^0 = 1$, $\beta_2^0 = 1.5$.

n error	β_1				β_2			
	case LS	a) LAD	case LS	b) LAD	case LS	a) LAD	case LS	b) LAD
$n = 300$								
$\mathcal{N}(0, 2)$	1	1	1	1	1.51	1.5	1.49	1.5
$\mathcal{L}(0, 2)$	0.99	1	0.98	0.99	1.52	1.5	1.53	1.5
$\mathcal{C}(0, 2)$	1.17	1	0.96	1	1.56	1.5	1.48	1.5
$n = 100$								
$\mathcal{N}(0, 2)$	0.97	0.96	0.99	0.94	1.45	1.5	1.48	1.51
$\mathcal{L}(0, 2)$	0.99	0.99	1	0.99	1.39	1.49	1.39	1.50
$\mathcal{C}(0, 2)$	1.18	1	1.29	1	1.81	1.5	1.62	1.5

Table 6

Coverage probabilities for linear model, $\pi(X) = 0.8 + 0.2|x - 1|$, if $|x - 1| \leq 1$ and 0.95 elsewhere, $n = 100, 20$, error Gaussian, Laplace, Cauchy.

ε	$n = 100$					$n = 20$				
	CP_{LS}	CP_{LAD}	\hat{CP}_{LS}	NCP_{LS}	NCP_{LAD}	CP_{LS}	CP_{LAD}	\hat{CP}_{LS}	NCP_{LS}	NCP_{LAD}
$\mathcal{N}(0, 1)$	0.95	0.943	0.954	0.945	0.802	0.967	0.921	0.955	0.994	0.785
$\mathcal{L}(0, 1)$	0.960	0.991	0.960	0.962	0.784	0.965	0.959	0.965	0.864	0.757
$\mathcal{C}(0, 1)$	0.985	1	0.986		0.745	0.984	0.999	0.979		0.737
$\mathcal{N}(0, 2)$	0.956	1	0.956	0.983	0.631	0.948	0.999	0.948	0.83	0.612
$\mathcal{L}(0, 2)$	0.952	1	0.952	0.889	0.665	0.967	0.999	0.965	0.993	0.588
$\mathcal{C}(0, 2)$	0.974	1	0.974		0.537	0.983	1	0.981		0.526

Table 7

Coverage probabilities for linear model, $\pi(X) = 0.8$, $n = 500, 20$, error Gaussian, Laplace, Cauchy.

ε	$n = 100$					$n = 20$				
	CP_{LS}	CP_{LAD}	\hat{CP}_{LS}	NCP_{LS}	NCP_{LAD}	CP_{LS}	CP_{LAD}	\hat{CP}_{LS}	NCP_{LS}	NCP_{LAD}
$\mathcal{N}(0, 1)$	0.954	0.948	0.951	0.964	0.800	0.961	0.913	0.943	0.947	0.848
$\mathcal{L}(0, 1)$	0.950	0.985	0.944	0.962	0.811	0.965	0.958	0.952	0.956	0.806
$\mathcal{C}(0, 1)$	0.983	1	0.978		0.769	0.992	0.999	0.978		0.656
$\mathcal{N}(0, 2)$	0.952	0.999	0.950	0.949	0.561	0.968	0.997	0.946	0.870	0.630
$\mathcal{L}(0, 2)$	0.955	1	0.952	0.958	0.583	0.964	0.998	0.946	0.993	0.620
$\mathcal{C}(0, 2)$	0.982	1	0.982		0.54	0.985	1	0.976		0.481

Table 8

Mean, standard deviation, median of parameter estimations for linear model, $\pi(X) = 0.8 + 0.2|x - 1|$, if $|x - 1| \leq 1$ and 0.95 elsewhere, $\beta^0 = 10$, error Gaussian, Laplace, Cauchy.

n	mean	mean	sd	sd	median	median
error	LS	LAD	LS	LAD	LS	LAD
$n = 100$						
$\mathcal{N}(0, 1)$	10	10	0.07	0.09	10	10
$\mathcal{L}(0, 1)$	10	10	0.10	0.08	10	10
$\mathcal{C}(0, 1)$	9.94	10	93	0.12	9.98	10
$\mathcal{N}(0, 2)$	10	10	0.15	0.18	10	10
$\mathcal{L}(0, 2)$	10	10	0.21	0.16	10	10
$\mathcal{C}(0, 2)$	10.7	10	59	0.25	9.92	10
$n = 20$						
$\mathcal{N}(0, 1)$	10	10	0.17	0.22	10	10
$\mathcal{L}(0, 1)$	10	10	0.24	0.22	10	10
$\mathcal{C}(0, 1)$	10	10	189	0.34	10	10
$\mathcal{N}(0, 2)$	10	10	0.34	0.42	10	10
$\mathcal{L}(0, 2)$	9.97	9.98	0.50	0.44	9.98	9.98
$\mathcal{C}(0, 2)$	8.97	10	29	0.65	10	10

Table 9

Mean, standard deviation, median of parameter estimations for linear model, $\pi(X) = 0.8$, $\beta^0 = 10$, error Gaussian, Laplace, Cauchy. True value $\beta^0 = 10$.

n	mean	mean	sd	sd	median	median
error	LS	LAD	LS	LAD	LS	LAD
$n = 100$						
$\mathcal{N}(0, 1)$	10	10	0.08	0.10	10	10
$\mathcal{L}(0, 1)$	10	10	0.11	0.09	10	10
$\mathcal{C}(0, 1)$	9.71	10	16	0.13	10	10
$\mathcal{N}(0, 2)$	10	10	0.17	0.21	10	9.99
$\mathcal{L}(0, 2)$	9.99	10	0.22	0.18	9.99	10
$\mathcal{C}(0, 2)$	13	10	1.84	0.26	9.97	9.99
$n = 20$						
$\mathcal{N}(0, 1)$	10	10	0.20	0.23	10	10
$\mathcal{L}(0, 1)$	10	10	0.26	0.25	10	10
$\mathcal{C}(0, 1)$	8.46	10	31.2	0.38	9.99	9.99
$\mathcal{N}(0, 2)$	9.99	9.98	0.37	0.47	9.98	9.98
$\mathcal{L}(0, 2)$	10	10	0.53	0.50	10	10
$\mathcal{C}(0, 2)$	9.22	10	34	0.82	10	10

Table 10

The acceptance rate of hypothesis $H_0 : \beta = \beta^0$ with respect to missing probability $1 - \pi(X)$.

	$1 - \pi(X) = 0.20$	$1 - \pi(X) = 0.50$	$1 - \pi(X) = 0.80$
complete data, LS method	1	0.996	0.976
complete data, LAD method	0.993	0.977	0.983
reconstituted data, LS method	1	0.992	0.874

Table 11

Empirical means, standard-deviations of differences between Y_i and its reconstituted (forecasted) value, with respect to estimation method and missing probability $1 - \pi(X)$.

	$1 - \pi(X) = 0.20$	$1 - \pi(X) = 0.50$	$1 - \pi(X) = 0.80$
$\text{mean}(Y_i - \hat{Y}_{i,0,LS})$	0.066	0.066	0.066
$\text{ean}(Y_i - \hat{Y}_{i,0,LAD})$	0.321	0.321	0.321
$\text{mean}(Y_i - \hat{Y}_{i,LS})$	0.002	-0.003	-0.0158
$\text{sd}(Y_i - \hat{Y}_{i,0,LS})$	3.35	3.35	3.35
$\text{sd}(Y_i - \hat{Y}_{i,0,LAD})$	3.36	3.36	3.36
$\text{sd}(Y_i - \hat{Y}_{i,LS})$	1.71	3.44	6.65